Interactive comment on "The impact of structural error on parameter constraint in a climate model" *by* D. McNeall et al.

Anonymous Referee #1

Received and published: 20 June 2016

Review of paper: "The impact of structural error on parameter constraint in a climate model" by Doug McNeall et al.

Thank you for inviting me to review this paper.

The paper is interesting and important as it addresses whether a component of a GCM can be calibrated for one part of the globe, but applied elsewhere. Climate models are heavily dependent on transferability of parameterisation of sub-model structure, and a knowledge of when this fails is important.

I can see the aim of the paper, and it will be useful to have in the literature. However there did seem to be a slightly excessive use of statistical terminology. That's fine if the statistics is of standard form, but that's not the case here as the methods utilised are more novel. Please ensure that the literature is cited sufficiently well that any part of

this paper can be understood by calling upon the appropriate referenced papers.

Below are some comments that the authors might like to consider for a revised manuscript:

Overall points

The title is possibly too general. The emphasis is on DGVM modelling of forests, not general overall issues of structure.

The Abstract needs to be something that can be read in isolation, such that the reader can obtain a strong idea what the paper is about. To my mind, there is some repetition (e.g. three times says this uses "a history matching approach", and yet doesn't define what this actually is). Removing repetition can make space for more details. Extra description of the parameters changed would be helpful, rather than a vague "parameters that lead to a realistic forest fraction".

Reviewing this, I'm trying to really understand what the main thrust of this paper is about, in the statistical/algorithm sense. Can I confirm that the over-arching message is that quantity delta in Eqn (1) is important, can be characterised, and shows geographical variation. To my mind, that is a powerful result. It basically says if (i) not enough process representation is introduced in to a model, then structure deficiency gets masked by parameter fitting, and (ii) doing so will create problems between different locations. It would be nice to acknowledge that structural errors presumably also reduce confidence in any model for future projections, even when just at a single region where it performs well for contemporary periods.

Page 9, starting "Does this region represent......". This is a critical part of the paper, discussing how in effect a standard best-fit might not always be appropriate. Can the discussion be led back to Eqn (1), and in particular the structural delta parameter? (Also line 1, page 9, I cannot see in a Table or diagram what the alternative potential values are, for comparison against the default inputs – apologies if I've missed some-

thing). Where are the local, or continent-scale, delta values given?

Details

P2, line 10. Again, please give the reader some idea what "History matching" is, given other quantities such as "calibration" and "tuning" are defined at this point.

Around lines P2, lines 25-29. It would be really nice to have more concrete reasons why emulators, parameterisations etc are needed. This usually comes down to two factors: (1), computational speed prevents very high resolution modelling, even if the processes are more fully understood. For example, parameterisation of convection. (2), we don't know what the values should be, and these may exhibit strong regional heterogeneity. The latter is more the case for this paper, with questions asked as to what are the appropriate number of plant functional types that should be in land surface models – and if the number is high, can for example EO provide the values.

Check notation is consistent throughout. P3, line 23, FAMOUS is described as a "climate simulator". In the minds of the authors, is this different to a standard GCMs. Do they regard FAMOUS's reduced resolution as removing it from being regarded as a full GCM?

Again, in Section 1.3, this is now the 7th or 8th time that "history matching" is mentioned – it would be good to help the reader as to what it is, even if it is only to provide a methodological citation at this point.

Cox (2001) is a technical note. Better to give a peer-reviewed reference?

P5, line 1. I don't understand the context of the sentence: "The Amazon region is not wet enough for a fully humid region to exist". If this refers to the FAMOUS model, and in particular its atmospheric response, then this will make any DGVM fail if rainfall totals are too small.

P5, discussion of beta parameter. In a similar vain to the comment above, is it OK to treat the atmospheric beta parameter as a "nuisance" parameter? Isn't there a risk that

C3

errors in GCM-projected precipitation – for example - will affect best-fit parameters in Table 1?

P5, line 18. From code that is shared with other centres, TRIFFID has a rapid spin-up option to near-equilibrium. Does it really need 10000 years?

Trivial thing, but it might be nice in Figure 2 to write as S.E.Asia (not SEASIA).

Can I confirm that a reader could find all details of the emulator in the Roustant et al 2012 paper. So, for instance, what a "leave-one-out cross validation metric" is.

Figure 7 I find very useful as it allows assessment of the geographical differences, providing more information that the global parameterisation Table 3. There are quite a few statistical methods available to determine parameter importance and/or nuisance parameters. An extra sentence stating what additional benefit the FAST algorithm brings would be helpful – i.e. beyond just the Saltelli reference.

Figure 8 is important as it shows how the Amazon has a difference response. Or put another way, a calibration of NL0 and V_CRIT_ALPHA for the Amazon could find a pair of parameters that would clearly be sub-optimal when applied to the other 3 regions. And vice-versa. I'd like to see more discussion around Figure 8, how it demonstrates the structural problems (i.e. very different responses to NL0 and V_CRIT_ALPHA, depending on location), and again – can this be related back to the delta parameter? This will also link better to the paper title, which is about model structural problems.

Figure 13 is nice and clear, and in many ways it is a shame that the paper is so long in technical details before getting to that point. Obviously this is a slightly naïve comment, but could it simply be that the trees of the Amazon have evolved differently to those of Africa. This could possibly be due to different imposed climatologies that the trees have adapted/acclimated to. So one conclusion of this paper could simply be that any land surface model such as TRIFFID requires a parameter mask, or ancillary fields, that are different for different places. The paper hints at this, page 16, in "Causes of discrep-

C4

ancy", where different rooting depths are considered. One future work extension might therefore be to include a root depth as a geographically-varying parameter, to add to those in Table 1? Would this then collapse delta down to zero for all locations?

C5